Jointly Learning Bilingual Sentiment and Semantic Representations for Cross-Language Sentiment Classification

Huiwei Zhou(✉), Yunlong Yang, Zhuang Liu, Yingyu Lin, Pengfei Zhu, and Degen Huang

School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, Liaoning, China
{zhouhuiwei,huangdg}@dlut.edu.cn,
{SDyyl_1949,zhuangliu1992,cdzpf}@mail.dlut.edu.cn,
lyydut@sina.com

Abstract. Cross-language sentiment classification (CLSC) aims at leveraging the semantic and sentiment knowledge in a resource-abundant language (source language) for sentiment classification in a resource-scarce language (target language). This paper proposes an approach to jointly learning bilingual semantic and sentiment representations (BSSR) for English-Chinese CLSC. First, two neural networks are adopted to learn sentence-level sentiment representations in English and Chinese views respectively, which are attached to all word semantic representations in the corresponding sentence to express the words in the certain sentiment context. Then, another two neural networks in two views are designed to jointly learn BSSR of the document from word representations concatenated with their sentence-level sentiment representations. The proposed approach could capture rich sentiment and semantic information in BSSR learning process. Experiments on NLP&CC 2013 CLSC dataset show that our approach is competitive with the state-of-the-art results.

Keywords: Jointly learning · Cross-language · Sentiment classification

1 Introduction

With the exponential growth of online reviews, sentiment classification has attracted much attention in the field of natural language processing (NLP). Lots of previous researches [1–3] focus on the sentiment classification in English, the most commonly studied language in NLP. Such researches produce a great quantity of high-quality labelled corpora and sentiment knowledge in English. However, sentiment resources are imbalanced in different languages. To leverage resources in a resource-rich language (such as English) for sentiment classification in a resource-scarce language (such as Chinese), cross-language sentiment classification (CLSC) approaches have been investigated.

Existing CLSC approaches concentrate on bridging the gap between the source and the target languages. Machine translation (MT) services are usually employed to connect the two languages by translating one language (view) to another language.

J. Wen et al. (Eds.): CCIR 2017, LNCS 10390, pp. 149–160, 2017.
https://doi.org/10.1007/978-3-319-68699-8_12

Then training and testing are performed in one view [4] or in two independent views respectively [5, 6]. However, learning language-specific classifiers in each view fails to capture the common sentiment and semantic information between two languages during training process. Besides, MT services unavoidably introduce translation errors, which may even change the sentiment polarities of translated texts [6].

With the revival of interest in deep learning, common semantic representations [7–9] for the source and the target languages are also used to eliminate the gap between the two languages. Common semantic representations project data from the two languages into a common space, where training and testing are performed. However, high-quality common semantic representations require a large-scale parallel corpus, which is not always readily available in many real-world scenarios. Meanwhile, though common semantic representations capture cross-language semantics between the two languages, sentiment knowledge is hard to be exploited. Recently, some researches incorporate sentiment knowledge into common semantic representations to enhance the ability of sentiment expression [10–13]. However, these approaches learn semantics and sentiment separately, and fail to adjust the complex relations between sentiment and semantics.

This paper proposes an approach to jointly learning bilingual semantic and sentiment representations (BSSR) for CLSC. First, MT services are used to translate English training data into Chinese and translate Chinese test data into English. Then, two convolutional neural networks are applied to learn the sentiment representation of each sentiment fragment in the two views, based on the polarities of sentiment words, respectively. The learned fragment sentiment representations of all sentiment words in a sentence are averaged up as the sentence-level sentiment representations (SLSR). Finally, two Long Short-Term Memory (LSTM) in two views are designed to learn BSSR of the documents from the concatenations of word semantic and sentence-level sentiment representations, since LSTM has shown powerful ability on modeling long sequences [3]. Our approach is motivated by the intuition that sentence-level sentiment representations are composed by fragment sentiment representations in the sentence, and the words in different sentences express different sentiment though their word semantic representations (WSR) are the same.

The proposed BSSR could capture sentiment and semantic information in cross-language document representations without relying on extra parallel corpora. Experiments on NLP&CC 2013 CLSC dataset show that our approach gets competitive performance compared with the state-of-the-art CLSC systems.

The major contributions of this work can be summarized as follows:

- We propose an approach to jointly learning BSSR of documents for CLSC, with only English training data and their translations. In the BSSR learning process, the sentiment information and cross-language semantic information can be adjusted to optimize CLSC performance.
- Cross-language semantic information is captured by minimizing the differences between the original and translated languages. It is proved that cross-language semantic information could bridge the gap between two languages effectively.

- We simply learn document representations from word semantic representations concatenated with sentence-level sentiment representations, which considers the words in the certain sentiment context. To the best of our knowledge, this representation approach has not yet been used for CLSC.

2 Related Work

In this section, we review the literature related to this paper from two perspectives: cross-language sentiment classification and sentiment classification.

2.1 Cross-Language Sentiment Classification (CLSC)

The goal of CLSC is to bridge the gap between the source language and target language. Wan [5] translates both the training and testing data to train different models in both source and target languages, which could learn language-specific classifiers in each view. However it is hard to capture the common sentiment and semantic knowledge of the two languages during training process.

Recently, there has been an interest in common semantic representations for two languages. Some approaches to common representation learning [7, 12, 13] require parallel data. These works use an encoder-decoder architecture with a shared hidden layer to learn common semantic representations. Zhou et al. [14] propose a cross-lingual representation learning model which simultaneously learns both the word and document representations in both languages. However, large-scale task-related parallel corpora may be a scarce resource for the English-Chinese sentiment classification task.

Some other works simply utilize the labelled corpora and their translations to learn common semantic representations. Zhou et al. [11] learn bilingual sentiment word embeddings (BSWE) with denoising autoencoders. BSWE is hard to adjust the complex relations between sentiment and semantics to optimize the CLSC performance. Zhou et al. [15] use a hierarchical attention model which is jointly trained with the bidirectional LSTM network to learn the document representations.

2.2 Sentiment Classification

Traditional sentiment classification approaches usually use a sentiment lexicon. The lexicon-based approaches compute the sentiment polarity for each text based on sentiment words and negatives [1]. Machine learning-based approaches employ sentiment words as important features to construct sentiment classifiers.

Tang et al. [2] incorporate sentiment polarity of the text to learn sentiment-specific word embeddings. Tang et al. [3] connect the target word representations with its context word representations to learn the semantic representations related to a target by LSTM models. Without requiring parser, CNN or LSTM could capture the fragment sentiment representations in a sentence by leveraging contexts of the sentiment words [16].

Inspired by Tang et al. [3], we attach contextual sentiment representations to the word semantic representations. Such concatenation representations are exploited for jointly learning BSSR in CLSC task.

3 Bilingual Semantic and Sentiment Representations for CLSC

Sentiment is expressed by phrases or sentences rather than by words [17]. The same words in different contexts could express different sentiment. To incorporate rich contextual sentiment information into BSSR learning process, we first extract fixed-length sentiment fragments centered at each sentiment word, and use CNN to learn the fragment sentiment representations based on the polarities of sentiment words. The fragment sentiment representations in corresponding sentences are averaged up as the SLSR (Sect. 3.1).

Then, the SLSR are attached to word semantic representations to express the words in the specific sentiment context. Finally, the concatenations are fed into LSTM for jointly learning BSSR.

3.1 Sentence-Level Sentiment Representations (SLSR) Learning

The context words within a window $[-2,2]$ centered at each sentiment word $sent_i$ are considered as its sentiment fragment: $\{word_{i-2}, word_{i-1}, sent_i, word_{i+1}, word_{i+2}\}$. We extract these sentiment fragments, and adopt two CNN to learn sentiment representations of each sentiment fragment in English and Chinese views respectively.

CNN for SLSR Learning. The fragment sentiment representations learned with CNN in the two languages follows the same process. Let $x_i \in \mathbb{R}^d$ be the d-dimensional word representations. Then the sentiment fragment could be represented as $x = \{x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}\}$ with the word representations. A convolution operation involving a filter $w \in \mathbb{R}^{h \times d}$ is applied to the sentiment fragment to produce a new feature. We set the filter window size $h = 3$ in our work. For example, a feature c_t is generated from a window of words $x_{t:t+2}$ by $c_t = f(w \cdot x_{t:t+2} + b)$, where $x_{t:t+2}$ refers to the concatenation of x_t, x_{t+1} and x_{t+2}. Here b is a bias term and f is a non-linear function such as the hyperbolic tangent. This filter is applied to each possible window of words in the sentiment fragment to produce a feature map $c = [c_{t-2}, c_{t-1}, c_t]$, with $c \in \mathbb{R}^3$.

We apply a max pooling operation over the feature map and take the maximum value $\hat{c} = \max\{c\}$ as the feature corresponding to this filter. 200 filters are used in our work to obtain 200 features. These features are fed to a multilayer perceptron to produce the fragment sentiment representations, which are passed to a logistic regression to predict the sentiment polarities.

After obtaining the fragment sentiment representations from the CNN model for each sentence, we average them to produce the SLSR.

We also introduce LSTM to learn fragment sentiment representations to compare with CNN. The same sentiment fragment used in CNN is fed into LSTM to learn the SLSR as well.

LSTM is designed to cope with the gradients vanishing or exploding problem of RNN [18]. It introduces a gating mechanism, which comprises four components: an input gate i_t, a forget gate f_t, an output gate o_t, and a memory cell c_t. For the standard LSTM, each of the three gates receives the information from the inputs at current time step and the outputs at previous time step. In our work, we adopt an LSTM variant, which adds the "peephole connections" to the architecture [19] to let the memory cell c_{t-1} directly control the gates as follows:

$$i_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + V^{(i)}c_{t-1} + b^{(i)}) \tag{1}$$

$$f_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1} + V^{(f)}c_{t-1} + b^{(f)}) \tag{2}$$

$$o_t = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1} + V^{(o)}c_t + b^{(o)}) \tag{3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W^{(c)}x_t + U^{(c)}h_{t-1} + b^{(c)}) \tag{4}$$

$$h_t = o_t \odot \tanh(c_t) \tag{5}$$

where W, U and V are the transition matrices for the input x_t, the hidden state vector h_{t-1}, and the memory cell c_{t-1}, respectively, b is a bias term of the hidden state vector, σ represents the sigmoid function, and $\odot$ denotes component-wise multiplication. In our work, hidden representations at every time step are fed into a mean pooling layer to obtain the final representations.

3.2 Jointly Learning Bilingual Sentiment and Semantic Representations (BSSR) of Documents

The gated recurrent architecture of LSTM is suitable for encoding sentiment and semantics of words and inherent relations of words into document representations. In this paper, we employ two LSTM for BSSR learning and the framework of jointly learning is shown in Fig. 1.

As shown in Fig. 1, one LSTM is used to produce the English representations of documents, while the other is used for the representations of their Chinese translations. These two LSTM are designed to minimize the difference between the representation pairs for capturing the cross-language semantic information. Meanwhile, the polarity labels of documents are used to guide the learning process for incorporating sentiment information into BSSR.

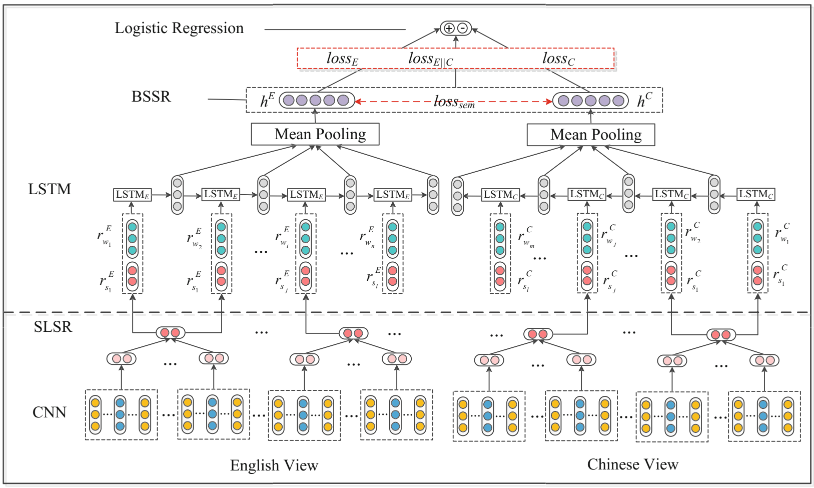

Fig. 1. The framework of jointly learning BSSR with LSTM.

For an English training document d_k^E containing p sentences and n words, it can be represented by a word sequence $\{w_1^E, w_2^E, \ldots, w_i^E, \ldots, w_n^E\}$. Once the j^{th} SLSR $r_{s_j} \in \mathbb{R}^{d_2}$ are attached to its WSR $r_{w_i} \in \mathbb{R}^{d_1}$. The input matrix $d_k^E \in \mathbb{R}^{(d_1+d_2)\times n}$ of the English view can be represented as:

$$d_k^E = \left\{ \begin{bmatrix} r_{w_1} \\ r_{s_1} \end{bmatrix}, \begin{bmatrix} r_{w_2} \\ r_{s_1} \end{bmatrix}, \ldots, \begin{bmatrix} r_{w_i} \\ r_{s_j} \end{bmatrix}, \ldots, \begin{bmatrix} r_{w_n} \\ r_{s_l} \end{bmatrix} \right\} \tag{6}$$

If no SLSR is found during the learning process, we attached a zero vector which dimension is also d_2. The same step repeats for its Chinese translation d_k^C. Thus paired documents (d_k^E, d_k^C) are expressed by the concatenations of WSR and their SLSR, which are fed to LSTM model in English (LSTM_E) and Chinese (LSTM_C) respectively.

As for an English training document d_k^E, LSTM_E recursively computes its representations h_k^E. The same step repeats for its Chinese translation d_k^C to generate document representations h_k^C. The learned document representations h_k^E, h_k^C and the concatenation of them $[h_k^E \,||\, h_k^C]$ are used to compute the polarity probabilities $p_E(s \,|\, d_k; \xi^E)$, $p_C(s \,|\, d_k; \xi^C)$ and $p_{E||C}(s \,|\, d_k; \xi^{E||C})$ through a logistic layer, where $s = \{0, 1\}$ is the sentiment polarity of a document, and ξ^E, ξ^C, and $\xi^{E||C}$ are the hyper-parameters of the models.

The sum of three errors predicted by h_k^E ($loss_E$), h_k^C ($loss_C$), and $[h_k^E \,||\, h_k^C]$ ($loss_{E||C}$) are used as the sentiment loss function to adjust the sentiment information in BSSR learning:

$$loss_{sen} = loss_E + loss_C + loss_{E||C} \tag{7}$$

Meanwhile, we also define a semantic loss function to minimize the differences between the representation pairs to adjust the cross-language semantic information:

$$loss_{sem} = \left\| h_k^E - h_k^C \right\|^2 \tag{8}$$

To investigate the relations between sentiment and cross-language semantic information, we learn BSSR by minimizing the loss function:

$$loss = \alpha \cdot loss_{sen} + (1 - \alpha) \cdot loss_{sem} \tag{9}$$

where α is the hyper-parameter which controls the weight of sentiment information.

4 Experiment

4.1 Experimental Settings

Dataset. The proposed approach is evaluated on NLP&CC 2013 CLSC dataset[1]. The dataset consists of product reviews in three domains: Book, DVD, and Music. Each domain contains 4,000 English labelled documents as training data and 4,000 Chinese unlabeled documents as test data.

Sentiment Words. Based on the Chi-square method, 2245 English sentiment words in MPQA subjectivity lexicon [20] and 3010 Chinese sentiment words in affective lexicon ontology [21] are used for SLSR learning. Note that if there are any negative word in the sentiment segment, the polarity of sentiment word is inversed. Here, we choose the negation words as follows: no, never, none, few, little, hardly and seldom. The Chinese negation words are listed as follows: 不 (no), 没有 (none), 很少 (few), 几乎不 (seldom), 从未 (never).

Tools. In our experiments, Google Translate[2], is adopted for both English-to-Chinese and Chinese-to-English translation. All corpora are segmented by Stanford Word Segmenter[3]. CNN and LSTM are developed based on Theano [22]. The SLSR are learned for 50 epochs, and BSSR are learned for 20 epochs. To make a comparison between SLSR and WSR, we use 200-dimensional WSR to produce 200-dimensional SLSR. The details of the SLSR learning are shown in Sect. 3.1. As for the BSSR, the model has 30-dimensional SLSR and 200-dimensional WSR. Besides, publicly available Global Vectors [23] (GloVe)[4] are adopted as the pre-trained WSR in English.

1 http://tcci.ccf.org.cn/conference/2013/dldoc/evsam03.zip.

2 http://translate.google.cn/.

3 http://nlp.stanford.edu/software/segmenter.shtml.

4 http://nlp.stanford.edu/projects/glove/.

We pre-train the Chinese WSR on a 27 MB-sized unlabeled dataset (also included the NLP&CC 2013 CLSC training set) with GloVe toolkit.

Evaluation Metric. The performance is evaluated by the correct classification accuracy of each domain, and the average accuracy of three domains, respectively. The domain accuracy is defined as:

$$Accuracy_n = \frac{\#system\_correct}{\#system\_total} \tag{10}$$

where n is one of the three domains, $\#system\_correct$ and $\#system\_total$ stand for the number of being correctly classified reviews and the number of total reviews in the domain n, respectively.

The average accuracy is shown as:

$$Average = \frac{1}{3}\sum_{n} Accuracy_n \tag{11}$$

4.2 Evaluations on BSSR

We compare our model with several different representations learning methods in Table 1. The following methods are used for comparison.

Table 1. The comparison of different representations

Methods	Book	DVD	Music	Average
CHN	74.20	74.02	73.50	73.91
WSR	77.45	77.45	77.00	77.30
SLSR	74.00	75.85	66.38	72.08
LSTM-BSSR	79.88	79.38	77.58	79.40
CNN-BSSR	81.78	83.00	80.68	**81.82**
CNN-SegBSSR	78.92	78.67	77.30	78.30

CHN: The labeled English reviews are translated to Chinese. Then training and testing are performed only in Chinese view with LSTM.

WSR: Only the 200-dimensional WSR is fed into the two LSTM models to classify the document.

SLSR: Only the 200-dimensional SLSR is fed into the multilayer perceptron model to classify the document. We use the SLSR learned by the CNN model.

CNN-BSSR: SLSR learned by the CNN are used in the BSSR learning. The hyper-parameter α is set to 0.5.

LSTM-BSSR: SLSR learned by the LSTM are used in the BSSR learning. The hyper-parameter α is set to 0.5.

CNN-SegBSSR: To evaluate our model, we only concatenate the SLSR with the sentiment words and its context words in the window of $[-2, 2]$. We use the SLSR learned by the CNN model. As for the other words out of the window, we concatenate a zero vector **0** instead of the SLSR.

From Table 1, we can see that:

- The bilingual representations (**WSR**, **CNN-BSSR** and **LSTM-BSSR**) outperform the sole Chinese representations (**CHN**).
- **SLSR** gets a lower accuracy than others. The reason is that only the coarse-grained sentence-level sentiment representations could not capture polarity of the document. WSR is still needed.
- Both **CNN-BSSR** and **LSTM-BSSR** outperform **WSR** in all the three domains. The average accuracy of **CNN-BSSR** reaches 81.82%, which is 4.52% higher than that of **WSR**. These indicates that **SLSR** and **WSR** are both effective for sentiment expression and their concatenation could improve the performance further.
- **CNN-BSSR** outperforms **LSTM-BSSR**, which indicates that CNN is superior in sentence-level sentiment representation.
- **CNN-SegBSSR** cannot catch up with **CNN-BSSR**, which demonstrates that SLSR is important for each word expression in sentences.

4.3 Influences of Sentiment and Cross-Language Semantic

We investigate the influences of sentiment and cross-language semantics for BSSR learning in Fig. 2. We use the CNN-BSSR with weighting parameter α varying from 0 to 1. From Fig. 2, we can see that:

- The three curves start from the initial accuracies of the sole cross-language semantic information, and then increase to their individual highest accuracy, finally fall to the individual accuracy of the sole sentiment information. These results indicate that both the sentiment and semantic information are effective in CLSC task and the combination of the two types of information outperforms either one of them obviously.

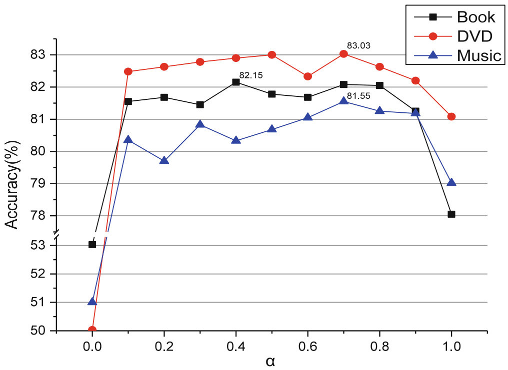

Fig. 2. Relations between sentiment and semantics with different α.

- When $\alpha = 0$, the model only has the cross-language semantic loss and produces poor results. When $\alpha = 1$, the model only has the sentiment loss and gets relatively better results than $\alpha = 0$.
- With the weighting parameter varying from 0 to 1, CLSC performances change in the Book, DVD, and Music domains. The best performances are obtained when the weighting parameter α is 0.4 (82.15% accuracy), 0.7 (83.03% accuracy), and 0.7 (81.55% accuracy) in Book, DVD and Music domains respectively. It shows that the effects of sentiment and cross-language semantics are different in various domains.

4.4 Comparison with Related Work

Table 2 shows the comparison results of our approach with some state-of-the-art systems on NLP&CC 2013 CLSC dataset.

Table 2. Cross-language sentiment classification accuracy of different methods.

	Book	DVD	Music	Average
Gui et al. [24]	80.50	82.20	79.70	80.80
Zhou et al. [11]	81.05	81.60	79.40	80.68
Zhou et al. [15]	82.10	**83.70**	81.30	82.40
Ours	**82.15**	83.03	**81.55**	82.24

Gui et al. [24] propose a transfer detection approach to learn sentiment classifiers with removing the noise from the transferred samples to avoid negative transfers (NTD) and achieved an 80.80% average accuracy. Such two-view approach is difficult to capture the common sentiment and semantics between two languages.

Zhou et al. [11] propose the bilingual sentiment word embedding algorithm based on denoising autoencoders. It learns common representations of two views by a semantic learning phase and a sentiment learning phase and achieves 80.68% accuracy. The two-phase approach is hard to adjust the relations between sentiment and semantics for different application domains.

Zhou et al. [15] propose a hierarchical attention mechanism with bidirectional LSTM for bilingual representation. They achieve 82.40% accuracy benefiting from the word-level attention model and the sentence-level attention model. However, training bidirectional LSTM models in both English and Chinese views are time consuming and high cost.

Our approach incorporates contextual sentiment information into word semantic representations, and learns the bilingual sentiment and semantic representations jointly. We achieve a high accuracy of 82.24% by simply employing the training data and their translations. Comparing with the Zhou et al. [15], our model is easy to train and has a comparable result. Considering the attention method and bidirectional architecture are effective for the CLSC task, we would like to explore that in our future work.

5 Conclusion

This paper proposes an approach to jointly learning bilingual sentiment and semantic representations in a unified framework. The learned representations are employed for calculating the sentiment polarities of documents. In the learning process, the predicted errors are used as the sentiment loss to capture the sentiment information. Meanwhile, the differences between the English and Chinese representation pairs are used as cross-language semantic loss to capture the cross-language semantic information. The experimental results demonstrate that sentiment and cross-language semantic information are both effective for CLSC, and the performance of CLSC in different domains could be optimized by adjusting sentiment and cross-language semantic loss. Attention-based methods show effectiveness in the sentiment classification task, and we leave it as future work to further boost the performance of CLSC.

Acknowledgements. This research is supported by Natural Science Foundation of China (No. 61272375).

References

1. Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of ACL, USA, pp. 417–424 (2002)
2. Tang, D.Y., Wei, F.R., Yang, N., Zhou, M., Liu, T., Qin, B.: Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceeding of ACL, USA, pp. 1555–1565 (2014)
3. Tang, D.Y., Qin, B., Feng, X.C., Liu, T.: Effective LSTMs for target-dependent sentiment classification. In: Proceeding of COLING, Japan, pp. 3298–3307 (2016)
4. Li, S., Wang, R., Liu, H., Huang, C.R.: Active learning for cross-lingual sentiment classification. In: Zhou, G., Li, J., Zhao, D., Feng, Y. (eds.) NLPCC 2013. Communications in Computer and Information Science, vol. 400, pp. 236–246. Springer, Heidelberg (2013). doi:10.1007/978-3-642-41644-6_22
5. Wan, X.J.: Co-training for cross-lingual sentiment classification. In: Proceedings of ACL, Singapore, pp. 235–243 (2009)
6. Chen, Q., Li, W.J., Lei, Y., Liu, X.L., He, Y.X.: Learning to adapt credible knowledge in cross-lingual sentiment analysis. In: Proceedings of ACL, China, pp. 419–429 (2015)
7. Chander, S., Lauly, S., Larochelle, H., et al.: An autoencoder approach to learning bilingual word representations. In: Proceeding of NIPS, Canada, pp. 1853–1861 (2014)
8. Yoshikawa, Y.Y., Iwata, T., Sawada, H., Yamada, T.: Cross-domain matching for bag-of-words data via kernel embeddings of latent distributions. In: Proceedings of NIPS, Canada, pp. 1405–1413 (2015)
9. Rajendran, J., Khapra, M., Chandar, S., Ravindran, B.: Bridge correlational neural networks for multilingual multimodal representation learning (2015). arXiv preprint arXiv:1510.03519
10. Tang, X.W., Wan, X.J.: Learning bilingual embedding model for cross-language sentiment classification. In: the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, pp. 134–141 (2014)
11. Zhou, H.W., Chen, L., Shi, F.L., Huang, D.G.: Learning bilingual sentiment word embeddings for cross-language sentiment classification. In: Proceedings of ACL, China, pp. 430–440 (2015)

12. Klementiev, A., Ivan, T., Bhattarai, B.: Inducing cross lingual distributed representations of words. In: Proceedings of COLING, Indiana, pp. 1459–1474 (2012)
13. Jain, S., Batra, S.: Cross-lingual sentiment analysis using modified BRAE. In: Proceedings of EMNLP, Portugal, pp. 159–168 (2015)
14. Zhou, X.J., Wan, X.J., Xiao, J.G.: Cross-lingual sentiment classification with bilingual document representation learning. In: Proceedings of ACL, Germany, pp. 1403–1412 (2016)
15. Zhou, X.J, Wan, X.J and Xiao, J.G.: Attention-based LSTM network for cross-lingual sentiment classification. In: Proceedings of EMNLP, USA, pp. 247–256 (2016)
16. Collobert, R., Weston, J., Bottou, L., et al.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**, 2493–2537 (2011)
17. Wang, X., Liu, Y.C., Sun, C.J., Wang, B.X., Wang, X.L.: Predicting polarities of tweets by composing word embeddings with long short-term memory. In: Proceedings of ACL, China, pp. 1343–1353 (2015)
18. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
19. Gers, F.A., Schmidhuber, J.: Recurrent nets that time and count. In: The IEEE-Inns-Enns International Joint Conference on Neural Networks, vol. 3, pp. 189–194 (2000)
20. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. Int. J. Comput. Appl. **7**(5), 347–354 (2005)
21. Xu, L.H., Lin, H.F., Pan, Y., Ren, H., Chen, J.M.: Constructing the affective lexicon ontology. J. China Soc. Sci. Tech. Inf. **27**(2), 180–185 (2008)
22. Bastien, F., Lamblin, P., Pascanu, R., et al.: Theano: new features and speed improvements (2012). arXiv preprint arXiv:1211.5590
23. Pennington, B., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: Proceedings of EMNLP, USA, pp. 1532–1543 (2014)
24. Gui, L., Lu, Q., Xu, R., Wei, Q., Cao, Y.: Improving transfer learning in cross lingual opinion analysis through negative transfer detection. In: Zhang, S., Wirsing, M., Zhang, Z. (eds.) KSEM 2015. LNCS, vol. 9403, pp. 394–406. Springer, Cham (2015). doi:10.1007/978-3-319-25159-2_36